基于多重因素的个性化学习推荐系统 *

匡　容，杨振国，刘文印†

(广东工业大学 计算机学院，广州 510006)

摘　要：为解决现有学习推荐算法中存在的忽略对学生知识点掌握情况的分析、不能将知识掌握程度概率化等问题，提出一种基于多重因素的学习推荐方法。该方法综合考虑知识点的综合权重、错误率和失分率多个因素构建知识点掌握概率模型，并应用所提出的策略实现一个在线的个性化学习推荐系统。系统评估上对 200 名高中生进行了一项调查，本系统推荐 top-8 知识点的准确率达到 91.2%，F1 达到 78.4%。系统调查的结果显示了提出策略的有效性和可靠性。

关键词：定性推荐；推荐系统；电子学习

中图分类号：TP301.6　　doi: 10.19734/j.issn.1001-3695.2018.07.0471

Multiple factors based personalized learning recommendation system

Kuang Rong, Yang Zhenguo, Liu Wenyin†

(*School of Computer Science &Technology Guangdong University of Technology*, *Guangzhou* 510006, *China*)

Abstract: In order to solve the problems existing in the learning recommendation algorithm that ignore the analysis of the students' knowledge points and can not probabilize the knowledge mastery, this paper proposed a recommendation method based on multiple factors. The method focused on the comprehensive weight of knowledge points, error rate and loss rate, and built a knowledge point mastery probability model, and applied the proposed strategy to implement an online personalized learning recommendation system . In terms of the systematic evaluation, through a survey of 200 high school students, the accuracy of the top-8 knowledge points recommended by our system achieves significant performance, Precision: 91.2%, and F1: 78.4%. The results of the systematic survey reflect the effectiveness and reliability of the proposed strategy.

Key words: qualitative recommendation; recommender system; e-learning

0　引言

考试最早起源于中国，是学校检测知识掌握情况的重要手段。在校中学生每年需要面对大量的考试（课堂测试，周考、月考等）。然而，在实际过程中考试检测的作用未得到充分的利用。在大多数中学，学校没有完善的成绩管理系统[1]，大多以纸质成绩单存储，学生的历史成绩难以查询或分析；一个教师要负责上百个学生，难以针对每个学生的测试做具体分析，对学生学习效果的评价缺乏深入地挖掘分析；大部分学生因自身知识和总结归纳能力有限难以展开深入探究，常停留在成绩分数。如何最大程度地发挥测试的作用，为学生明确知识漏洞是本文关注的重点。

随着互联网技术的不断发展[2]，教育领域涌现出很多前沿技术与应用，如数据挖掘技术[3]、推荐系统技术[4]、慕课[5]、网易公开课[6]等，这些成果改变了传统教学模式。现有学习推荐工作大多采用认知诊断模型[7,8]引入习题知识点矩阵、协同过滤方法[9,10]计算相似学生答题记录之间的相似度，根据相似学生进行推荐或概率矩阵分解方法[11]将学生得分矩阵分解成一组潜在因子，再预测目标学生在习题上的得分进行推荐。然而，认知诊断方法不能将知识掌握程度概率化；协同过滤方法和概率矩阵分解方法均忽略了对学生知识点掌握情况的分析[12]，均导致推荐效果不佳。因此，针对目前中学的教育现状，及以上算法和技术的不足，本文运用基于知识点综合权重、错误率和失分率的多重因素来构建学生知识点掌握概率模型，从而确定为学生推荐的知识点。该概率模型还与传统的认知诊断方法、协同过滤方法和概率矩阵分解方法进行了大量实验对比，并在个性化学习推荐系统得到实现。本文实现的个性化学习推荐系统应用于湖南省湘潭县一中高三年级数学知识点的分析中，,促使学校更好地开展教学工作,提高教学质量[13]。

1　问题描述

在众多课程中，知识点的掌握程度是衡量学生知识掌握情况的最直接的标准，知识点能反映学生在学习过程中的不足和知识漏洞[14]。因此，知识点对于学生进行针对性学习具有重要

收稿日期：2018-07-27；**修回日期：**2018-09-19　**基金项目：**国家自然科学基金（61703109）；广东创新研究团队计划（2014ZT05G157）

作者简介：匡容（1994-），女，湖南湘潭人，硕士，主要研究方向为数据挖掘；杨振国，男，博士后，主要研究方向为多媒体、社交媒体、迁移学习；刘文印，男，教授，博导，主要研究方向为区块链、网络身份安全、假冒网站检测、机器视觉、图形识别、文本挖掘（liuwy@gdut.edu.cn）.

意义。在教学过程中试题是非常重要的教育资源。通过测试能知道学生是否掌握了某个知识点，根据学生的测试得分了解学生弱项知识点的情况；并通过进行试题分析明确测试要点。所以，从学生日常测试信息中，提取有效因素构建推荐模型，向学生进行知识点推荐，以期望帮助学生在有限的时间内，弥补最重要的知识漏洞，有效地提高考试表现。

表 1　知识点推荐涉及符号及描述

Table 1　Notations

符号	描述
S	学生集合
S_u	学生 u
T	试题集合
T_v	试题 v
K	知识点集合
K_n	知识点 n
G	学生成绩矩阵
$g_{S_uT_v}$	学生 S_u 在试题 T_v 的得分情况
Q	试题知识点关联矩阵
$q_{T_vK_n}$	试题 T_v 考查考查知识点 K_n 的情况
$e_{S_uT_v}$	学生 S_u 在试题 T_v 的错误率
$l_{S_uT_v}$	学生 S_u 在试题 T_v 的失分率
w_n	知识点 n 的综合权重
$Rec_{S_uK_n}$	系统为 S_u 推荐知识点 K_n 的指数

在知识点推荐问题中，有 U 个学生 $S=\{S_1, S_2, \cdots, S_u\}$ 参加某课程 V 道试题 $T=\{T_1, T_2, \cdots, T_v\}$ 的测试，试题 T_i 的标准分为 g_{Ti}，其中考查了 N 个知识点 $K=\{K_1, K_2, \cdots, K_n\}$，$U$ 个学生的习题得分矩阵 $G=[g_{SuTv}]_{U\times V}$，其中，$g_{SuTv}=a(0\leqslant a\leqslant=g_{Tv}$,$g_{Tv}$ 表示该试题 T_v 的标准分)表示学生 S_u 在试题 T_v 的得分为 a。试题知识点关联矩阵 $Q=[q_{TvKn}]_{V\times N}$，其中，$q_{TvKn}=0$ 表示试题 v 没有考查知识点 n，$q_{TvKn}=1$ 表示试题 T_v 考查知识点 K_n。由此，通过学生成绩得分 g_{SuTv} 及试题知识点关系 q_{TvKn} 提取学生 S_u 在知识点 K_n 上的失分率(*loss rate*, l_{SuKn}),错误率(*error rate*, e_{SuKn})以及该知识点的综合权重(*comprehensive weight*, w_n)，构建知识点推荐模型为学生进行知识点推荐，以期望帮助学生更加准确地了解自己的知识漏洞，及时弥补。文中涉及符号及相关描述如表 1 所示。

基于多要素的个性化知识点推荐算法基本思路主要由数据输入，多要素获取和输出推荐列表三个部分组成，如图 1 所示。每个部分的内容如下：

a）数据输入。通过对学生成绩、试题信息进行数据处理，获得测试知识点列表（用于标记试题）；并将获得的学生考试得分数据如表 2 所示，以及试题知识点关联数据如表 3 所示，将表中数据分别转换为习题得分矩阵 G 和知识点试题得分矩阵 Q 输入。

b）多要素获取。依据输入数据获取学生 S_u 在知识点 K_n 的错误率(e_{SuKn})、失分率(l_{SuKn})及知识点的综合权重(w_n)，对学生知识点掌握情况进行建模，获得知识点掌握概率模型，用推荐指数 Rec_{SuKn} 表示系统对学生 S_u 在知识点 K_n 上的推荐度，如表 4 所示。知识点推荐指数越高，越值得被推荐。

表 2　学生考试得分数据

Table 2　Instances of test scores

学生成绩								
	T_1	T_2	…	T_{j-1}	T_j	T_{j+1}	…	T_v
	5	5		5	20	12		10
S_1	5	5	…	0	12	5	…	7
S_2	0	5	…	5	10	9	…	6
…	…	…	…	…	…	…	…	…
S_u	5	0	…	0	5	6	…	9

表 3　试题知识点关联数据

Table 3　Relation matrix of questions and knowledge points

试题知识点关系				
	K_1	K_2	…	K_n
T_1	1	0	…	0
T_2	1	0	…	0
…	…	…	…	…
T_n	0	1	…	0

表 4　知识点推荐指数数据

Table 4　Recommendation result matrix

知识点推荐程度				
	K_1	K_2	…	K_n
S_1	0.723	0.511	…	0.234
S_2	0.152	0.415	…	0.267
…	…	…	…	…
S_u	0.347	0.871	…	0.454

c)输出推荐列表。根据学生知识点掌握程度，获得学生知识点当前的知识漏洞，并根据知识点薄弱程度（推荐指数 Rec_{SuKn}）进行排序，向每个学生进行 top-N 个性化知识点推荐。

2　知识点掌握概率模型

在个性化学习推荐系统中，学生知识点掌握概率模型的构建是重点。学生知识点掌握概率反映了学生对知识点的掌握情况。主要目的是根据学生需要，识别出该科目知

识点的考核要点及学生错误多的知识点和丢分多的知识点，并为其推荐一定知识相关的学习内容。其中的主要策略是根据知识点的综合权重、错误率、失分率三个重要影响因素构建学生知识点掌握概率模型，来获得学生对知识点的掌握情况。

Torre 教授提出的认知诊断模型（DINA 模型）[15]是在获取学生知识点掌握程度中使用最多的。DINA 模型用期望最大似然(expectation maximization,EM)[16]估算学生对试题参数猜测率(*slip*)和失误率(*guessing*)，然后用后验概率获得学生的答题情况和知识点掌握情况[12]。但 DINA 模型得到的知识点掌握情况只有掌握或未掌握两种，容易造成数据丢失，导致知识点推荐不

太准确；此外，DINA 模型求解过程中知识点不能过多，否则时间复杂度很高。本文针对 DINA 模型存在的缺陷，提出一种基于多因素的知识点掌握概率模型来求解学生知识点的掌握情况。认知诊断模型和学生知识点掌握概率模型的对比如图 2 所示，从图中的（3）可以看出学生 S_1 在认知诊断模型中已经掌握知识点 K_1，不会向学生 S_1 推荐知识点 K_1。而本文提出的模型求解结果认为学生 S_1 掌握还比较薄弱，在向学生推荐的知识点中会出现知识点 K_1。下面对学生知识点掌握概率模型中的各要素进行详细说明。

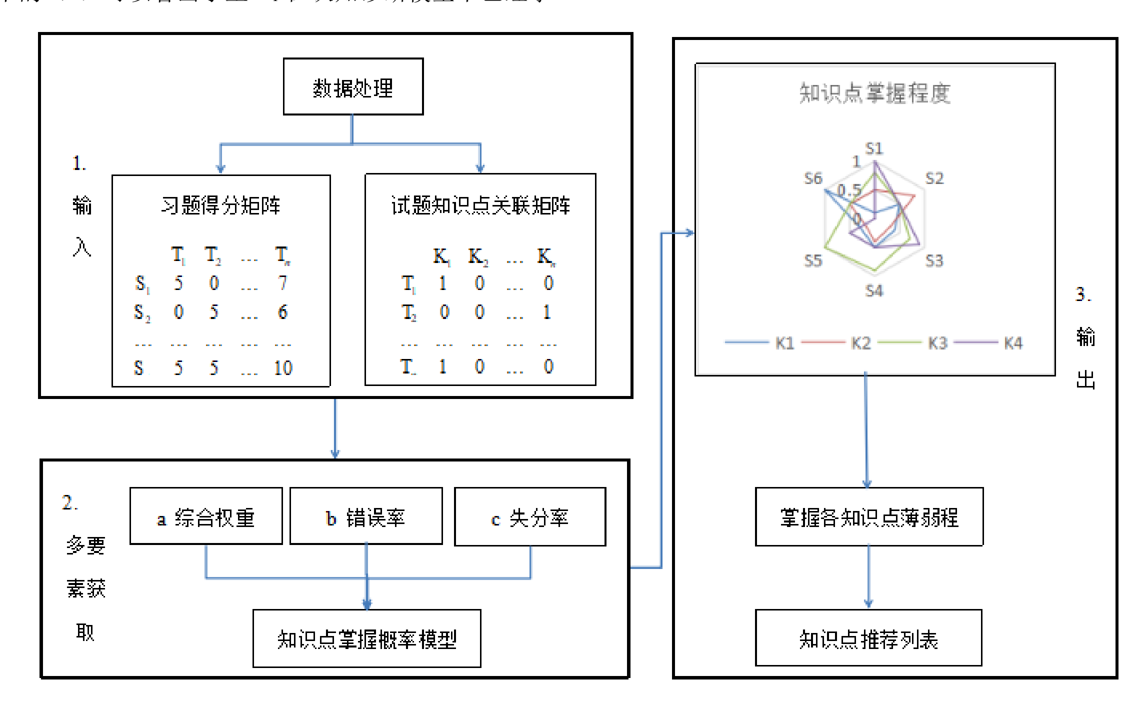


图 1　个性化知识点推荐算法的基本思路

Fig.1　Overview of the framework

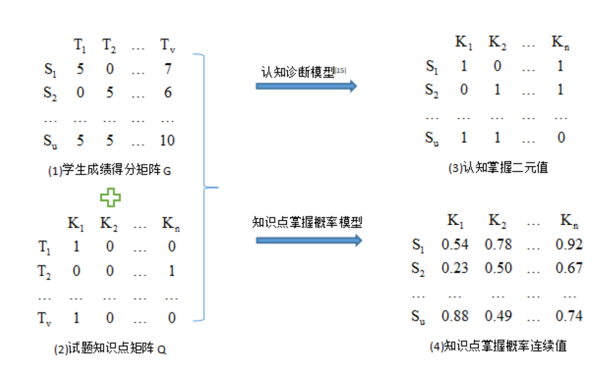


图 2　认知诊断模型和学生知识点掌握概率模型

Fig.2　Cognitive diagnosis model versus the proposed model

2.1　综合权重

知识点在考试查考过程中有一定的规律，用知识点的综合权重表示。综合权重表示知识点在考试中的重要程度，是知识点在多次考试中考核总分与多次试卷总分的比。某知识点的权重越高，说明该知识点越重要，是考试要点，值得被推荐。为了保证综合权重的合理性，在个性化学习推荐系统中，综合考虑知识点的历史考核权重及测试考核权重两个方面。历史考核权重为知识点在最近五年高考真题中的考核权重。万变不离其宗，真题从一定程度上反映了知识点的分布和考试的着重点。因此，分析知识点的历史权重至关重要。测试权重为知识点在学校测试（月考、期中考、期末考、模拟考等）中考核权重。教育不断发展，考试也日渐合理规范，新一年的考试对比往年，会有细微的调整，调整部分会通过学校测试反映出来。所以，分析知识点的测试权重也非常关键。

$$w_n' = \frac{\sum_{i=0}^{v} g_{T_i} q_{T_i K_n}}{\sum_{i=0}^{v}\sum_{j=0}^{n} g_{T_i} q_{T_i K_n}}, w_n' \in [0,1] \tag{1}$$

$$\begin{aligned} w_n &= a * his_w_n' + b * test_w_n' \\ w_n &= 0 (test_w_n' = 0) \end{aligned} \tag{2}$$

本文通过分析习题 T_v 考核知识点 K_n 的情况，获得知识点权重 w_n'，知识点权重 w_n' 计算公式如式(1)所示，g_{Ti} 表示试题 i 的标准分数，q_{TiKj} 表示试题 i 考查考查知识点 j 的情况，权重的取值范围为[0,1]。根据式(1)，通过分析多年高考真题中知识点考核情况，计算出在高考真题中知识点 n 的历史权重 his_w_n'，通过分析多次测试中知识点考核情况，计算出相应知识点的测试权重 $test_w_n'$。

综合权重 w_n 的计算如式(2)所示。因考试出题大多数参考历史真题，并根据最新的考纲进行部分调整，在日常测试中体现出来。因此，综合权重涉及历史权重与测试权重两部分，并设置影响因子 a，b 的比重进行线性相关，并通过分析 5 年高考真题和 4 次测试试卷获得影响因子 a：b 约为 7：3。若多次日常测试的知识考核权重为 0，表示该知识点考核情况进行了调整，在接下来的考试中将不再进行考核，则综合权重为 0。综合权重的取值范围为[0,1]。综合权重越高，说明该知识点越重要 。

2.2 错误率

错误率表示学生在多次考试中某知识点错误次数与该知识点考核次数的比。在个性化学习推荐系统中，某知识点未获得试题标准分即认定其错误。某知识点错误次数越多，说明该知识点是当前学生需要加强的点，也越值得被推荐。

$$e_{S_uK_n}=\frac{\sum_{k=0}^{v}p_k}{\sum_{i=0}^{v}q_{T_iK_n}},e_{S_uK_n}\in[0,1],$$
$$p_k=\begin{cases}0,p(g_{T_i}q_{T_iK_n}\neq g_{S_uT_i}q_{T_iK_n})\\1,p(g_{T_i}q_{T_iK_n}=g_{S_uT_i}q_{T_iK_n})\end{cases}. \tag{3}$$

其中:e_{SuKn}表示学生 S_u 在知识点 K_n 上的错误率，为学生 S_u 在知识点 K_n 错误次数与知识点考核次数的比，错误率 e_{SuKn} 的取值范围为[0,1]。通过从学生测试成绩及试题知识点关联关系获得学生在知识点 K_n 上的错误次数及知识点考核的总次数求得。未获得该试题满分的情况视为该试题错误，即相应考核知识点错误。错误率的计算如式(3)所示。错误率 e_{SuKn} 越高，说明学生 S_u 在知识点 K_n 考查中出错频率越高。

2.3 失分率

失分率表示学生在多次考试中某知识点丢失分数与该知识点考核总分数的比。失分率越高，说明学生该知识点掌握越不牢固，越影响考试的最终得分。所以，学生应该加强该知识点的学习。失分率与错误率有一定的联系，但非线性相关。在个性化学习推荐系统中，错误率越高的知识点失分率越高，但因主观题的得分情况除满分和零分外，可部分得分，两者非线性相关，所以失分率与错误率为两个不同的影响因素。

$$l_{S_uK_n}=\frac{\sum_{i=0}^{v}g_{T_i}-g_{S_uT_i}}{\sum_{i=0}^{v}g_{T_i}q_{T_iK_n}},l_{S_uK_n}\in[0,1] \tag{4}$$

失分率的计算如式(4)所示。l_{SuKn} 表示某考核知识点的失分率，其取值范围为[0,1]。通过从学生 S_u 在试题 T_i 的得分转换为学生 S_u 在知识点 K_n 的得分来计算，g_{Ti} 表示试题的标准分，g_{SuTi} 表示学生在试题 T_i 的真实得分，其取值范围为[0,g_{Ti}]，q_{TiKn} 表示知识点 K_n 与试题 T_i 的关联关系（考查或未考查）。失分率 l_{SuKn} 越高，说明学生 S_u 在考查知识点 T_i 过程中丢分越多，从一定程度上反映了学生的知识漏洞。

2.4 知识点掌握概率模型

个性化学习推荐系统的学生知识点掌握概率模型，综合考虑考虑综合权重，错误率和失分率多个因素三个重要因素，并根据这三个因素的构建推荐指数 Rec_{SuKn} 来表示学生知识点掌握概率，如式(5)所示。推荐指数越高，说明该知识点越值得被推荐。在实验过程中，设计三种方案 $Rec_{SuKn}1$，$Rec_{SuKn}2$，$Rec_{SuKn}3$，本文运用以下三种方案进行实验，并通过实验评测确认最终推荐方案应用于在线推荐系统中。在确认推荐知识点时，本文对推荐指数进行排序，为学生推荐其弱项知识点。系统根据推荐指数由高到低对知识点进行排序。推荐指数高的，优先推荐。根据学生将对推荐结果进行反馈，本文对推荐模型的推荐数量进行调整，力求推荐知识点更加合理。

$$\begin{aligned}&Rec_{S_uK_n}=f(w_n,e_{S_uK_n},l_{S_uK_n}),\\&Rec_{S_uK_n}1=w_n\frac{e_{S_uK_n}+l_{S_uK_n}}{2},\\&Rec_{S_uK_n}2=\frac{w_n+e_{S_uK_n}+l_{S_uK_n}}{3},\\&Rec_{S_uK_n}3=w_ne_{S_uK_n}l_{S_uK_n}.\end{aligned} \tag{5}$$

3 系统开发、应用及实验测评

3.1 数据集

为了验证学生知识点概率模型的有效性，本文在 Dataset 数据集，实验数据集如表 5 所示。Dataset 数据集是湘潭县一中高三学生连续多次在校真实的数学考试成绩，包含 1340 名学生，92 个试题，28 个知识点。其中，1340 名学生在 115 个具有试题得分构成习题得分矩阵 G；由 92 个试题和 28 个知识点（各知识点相互独立）构成试题知识点矩阵 Q，用 1 和 0 表示试题考查或未考查相应知识点。将通过数据处理后的数据作为系统输入。

表 5 Dataset 实验数据集

Table 5 Dataset

项目	数量
学生数量	1340
试题个数	92
知识点个数	28

3.2 实验及评测

为了验证算法的有效性，本文分别采用认知诊断模型(DINA)、协同过滤(CF)方法、概率矩阵分解(PMF)方法和本文提出的 Rec_{SuKn} 三种方案在湘潭县一中数学测试集上进行实验，并采用随机抽样对湘潭县一中 200 名学生展开以上方案推荐的 Top-8 的知识点列表准确度的问卷调查。

DINA 方法根据 DINA 诊断学生知识掌握程度，从推荐试题集合中选取学生掌握比较弱的知识点进行推荐；CF 方法通过学生已经有的试题得分，计算出学生之间的 Jaccard 相似度找到与之答题情况最相似的学生，再根据最相似学生的得分预测学生的得分，再根据预测结果得分的高低进行推荐；PMF 方法通过已有试题得分情况分解得到学生和试题的低维潜在因子向量，预测学生得分情况进行个性化推荐[8,12]；Rec_{SuKn} 方法从学生答题情况获取知识点综合权重，错误率和失分率多个重要因素构建知识点掌握概率模型，根据学生知识点掌握程度进行个性化推荐。

$$\begin{aligned}&Precision=\frac{TP}{TP+FP},\\&Recall=\frac{TP}{TP+FN},\\&F1=\frac{2*Precision*Recall}{Precision+Recall}.\end{aligned} \tag{6}$$

本文根据学生和教师对个性化学习推荐系统推荐的知识点结果的反馈，用准确率(*precision*)，召回率(*recall*)和 F_1 指标及运行时间(*running time*)来评价推荐效果，各指标的计算公式如式(6)所示。*Precision* 为系统推荐知识点正确个数与推荐推荐总数的比，*TP* 表示系统推荐知识点正确个数，*FP* 表示系统推荐知识点错误数量。准确率越高，说明系统为学生推荐的知识点越符合学生的实际学习情况。*Recall* 为系统推荐知识点正确个数与系统推荐个数和用户反馈未推荐的知识点总数的比,*FN* 表示用户反馈系统未推荐的知识点的数量。召回率越高说明系统为学生推荐的知识点越全面。F_1 值是精确率和召回率的调和均值。*Running time* 是算法在数据集上的运行时间。

不同推荐算法下推荐效果如表 6 所示。

表 6　不同推荐算法下推荐效果

Table 6　Compared with different algorithms

算法	precision	recall	F1	running time
DINA	0.498	0.467	0.482	4.47s
CF	0.553	0.541	0.547	4.18s
PMF	0.617	0.522	0.566	3.24s
$Rec_{SuKn}1$	0.672	0.596	0.632	2.71s
$Rec_{SuKn}2$	0.596	0.548	0.571	2.60s
$Rec_{SuKn}3$	0.912	0.688	0.784	2.46s

从表中数据对比可以看出，本文提出的基于多因素的个性化知识点推荐算法 Rec_{SuKn} 相对传统的 DINA 模型、CF 方法具有较好的效果。$Rec_{SuKn}1$，$Rec_{SuKn}2$，$Rec_{SuKn}3$ 算法的推荐效果相对传统的 DINA 模型最为明显，准确率分别提高了达到了 17.4%，9.8%，41.4%；*F*1 值分别提高了 15%，8.9%，30.2%。在本文提出的三种算法中 $Rec_{SuKn}3$ 的推荐效果最佳，准确率达到了 91.2%，*F*1 值为 78.4%，比 DINA 模型、PMF 方法、CF 方法准确率分别提高 41.4%，29.5%，35.9%，*F*1 高 30.2%，23.7%，21.8%。而且 $Rec_{SuKn}3$ 复杂度较低，运行时间最短，与传统 DINA 方法相比快 2.01s。主要原因在于：DINA 模型反馈学生对知识点的掌握情况只有掌握和未掌握两种，对知识点的掌握情况不能进行概率化，容易造成数据丢失，推荐效果欠佳,且数据集知识点个数较多,求解学生知识点掌握情况的过程时间复杂度较高；CF 方法根据相似学生的共性进行推荐，忽略了学生答题和知识掌握个性化，导致推荐效果欠佳；PMF 方法基于习题进行推荐，忽略了学生对知识点掌握情况的分析，导致其推荐准确率有待提高；而文中所提出的基于多要素的推荐方法，更能反馈学生对知识点的掌握情况，推荐结果更加贴合学生实际学习需要。

因此，选择 $Rec_{SuKn}3$ 的推荐策略应用与个性化学习推荐系统中给学生进行知识点推荐，并在实际应用过程中将推荐指数扩大一千倍，让用户能够清楚地了解每位学生各个知识点的掌握程度的差距，并激发学生发挥主观能动性，加强学习，提高考试表现。

在实验评估过程中，因学生未意识到自己的弱项知识点或者对知识体系不熟悉，可能反馈的未推荐知识点不全面，导致召回率不太准确。但从实验评估的整体考虑，学生对所有知识点的反馈均具有此特点，所以召回率的评估从一定程度上能说明以上情况。

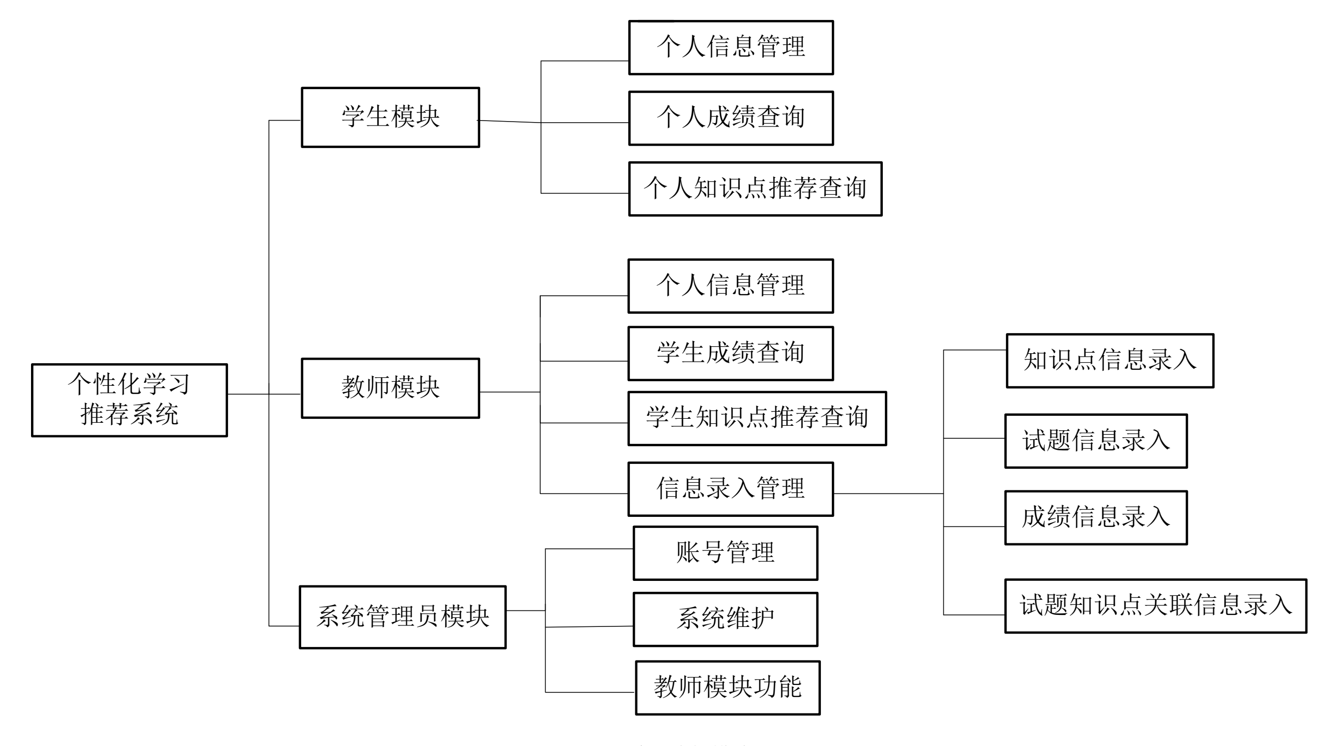


图 3　系统功能模块图

Fig.3　Relation matrix of questions and knowledge points

3.3　系统快照

个性化推荐系统是在 Linux[17]环境下用 DB2 作为数据库，采用 Python[18]语言进行编写的，系统使用 RazorSQL 进行数据库查询、SQL 的编辑、数据库管理；同时采用 Xshell 关联，Xftp

传输文件的方法，快速地对服务器进行上传、下载文件。

表 7　学生“丁昱夫”反馈信息统计表

Table 7　Feedback from a student “Ding Yufu”

推荐编号	1	2	3	4	5	6	7	8
正确/错误	√	×	√	√	√	√	√	√
补充个数	7	7	6	5	4	3	2	1

表 8　学生“丁昱夫”推荐效果评测表

Table 8　Performance example of “Ding Yufu”

推荐个数	1	2	3	4	5	6	7	8
准确率	1	0.5	0.667	0.75	0.8	0.833	0.857	0.875
召回率	0.143	0.111	0.222	0.333	0.444	0.556	0.667	0.889
F1	0.250	0.182	0.333	0.461	0.571	0.667	0.750	0.882

个性化学习推荐系统基于学生日常在校考试的真实数据，为用户提供个性化知识点推荐的浏览和学习服务，增强用户体验，提高网站的粘性。在系统中，根据用户角色（学生、教师、系统管理员）的不同设计不同的功能，系统功能模块图如图 3 所示。除了为用户推荐知识点外，本文还实现了信息管理，成绩查询，数据导入等功能，个性化学习推荐系统快照如图 4~7 所示。

从图 7 中可以看到，系统为学生“丁昱夫”进行 top-8 知识点推荐的结果，推荐知识点根据推荐指数排序，使学生能清晰明了地知道自己各个知识点的掌握程度。学生“丁昱夫”对推荐结果的反馈如表 7 所示，从该学生的反馈本文可知系统推荐的第 2 个知识点非丁昱夫的弱项知识点，其他知识点均推荐正确。“丁昱夫”在系统推荐 1 个知识点时，列出了 7 个未推荐的知识点。随着推荐知识点增多，学生补充知识点逐渐渐少。推荐第 8 个时，反馈中补充 1 个知识点。由此分析获得系统为其推荐的准确率和召回率及 $F1$ 指标如表 8 所示，因学生反馈系统推荐第 2 个知识点为选做题，并非不会作答，导致该未选做试题的相应知识点错误率和失分率很高，容易导致此处推荐结果不准确。但系统推荐准确度从第三个知识点开始不断升高。

据大部分学生反馈推荐知识点均为自己的知识漏洞，并根据推荐指数进行知识点的加强复习，能帮助学生有计划地进行巩固学习。随着系统推荐知识点增多时，学生认为系统推荐基本满足需求，且因部分知识点遗忘或记忆不清楚，所以未再补充系统未给自己推荐的知识点弱点。因此，后期完善，本文可考虑“题型”因素改进推荐模型。

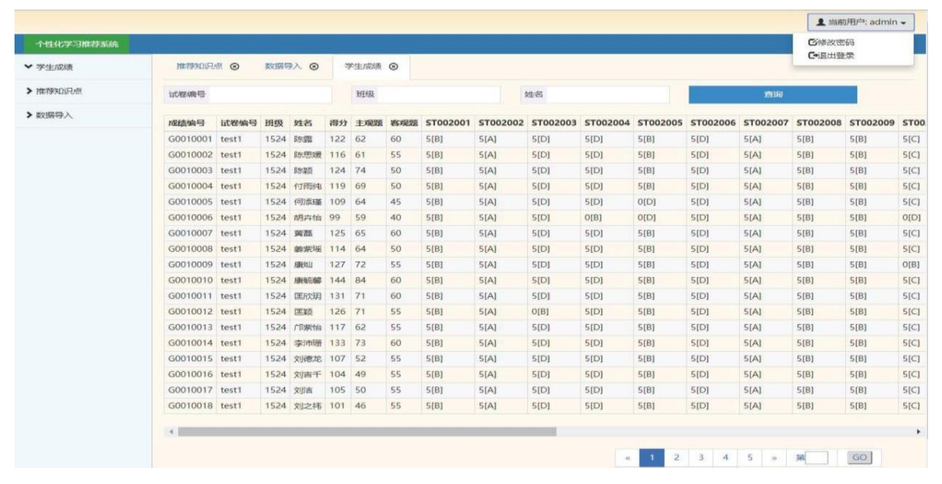


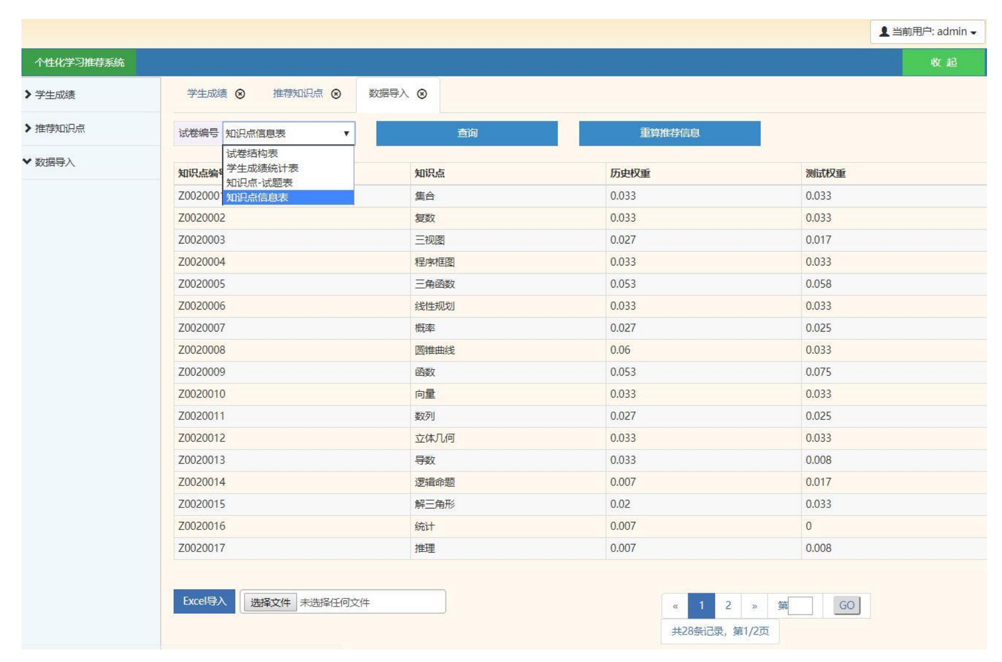
图 4　成绩查询

Fig.4　Results enquiry



图 5　知识点推荐

Fig.5　Knowledge point recommendation



知识点编号	知识点	历史权重	测试权重
Z0020001	集合	0.033	0.033
Z0020002	复数	0.033	0.033
Z0020003	三视图	0.027	0.017
Z0020004	程序框图	0.033	0.033
Z0020005	三角函数	0.053	0.058
Z0020006	线性规划	0.033	0.033
Z0020007	概率	0.027	0.025
Z0020008	圆锥曲线	0.06	0.033
Z0020009	函数	0.053	0.075
Z0020010	向量	0.033	0.033
Z0020011	数列	0.027	0.025
Z0020012	立体几何	0.033	0.033
Z0020013	导数	0.033	0.008
Z0020014	逻辑命题	0.007	0.017
Z0020015	解三角形	0.02	0.033
Z0020016	统计	0.007	0
Z0020017	推理	0.007	0.008

图 6　数据导入

Fig.6　Data import

学生成绩　数据导入　推荐知识点

班级　　姓名 丁昱夫　　推荐数量 8　　查询

姓名	班级	知识点	错误次数	考核次数	失分总数	失分率	推荐指数
丁昱夫	1523	导数与函数1	3	3	32	0.89	71.2
丁昱夫	1523	不等式选讲1（选做）	2	2	20	1	67
丁昱夫	1523	解析几何1	3	3	31	0.86	55.04
丁昱夫	1523	坐标系与参数方程1（选做）	3	3	15	0.5	33.5
丁昱夫	1523	导数	1	1	5	1	33
丁昱夫	1523	线性规划	3	3	15	1	33
丁昱夫	1523	数列1	2	2	11	0.46	29.44
丁昱夫	1523	三视图	1	1	5	1	27

共8条记录，第1/1页

图 7　知识点推荐结果

Fig.7　Predicted results

4　结束语

个性化学习推荐系统是针对普通中学生用户群的学习教辅系统，运用知识点综合权重和学生在相应知识点上的错误率、失分率多重因素，构建了多个学生知识点掌握概率模型，并与 DINA 方法、CF 方法、PMF 方法进行大量的对比实验，通过湘潭县一中学生实际应用评价，选择表现最佳的模型 $Rec_{SuKn}3$ 应用到系统中，系统默认推荐为 TOP-8 的知识点，实验评估准确率达到 91.2%，F1 达到了 78.4%，比 DINA 方法、CF 方法、PMF 方法准确率分别高 41.4%，29.5%，35.9%，显示了提出策略的有效性和可靠性。在一定程度上，系统能帮助中高级中学解决学校教辅系统不完善，教师一对一交流时间不够，学生分析问题能力不足等问题，明确学生知识漏洞，发挥主观能动性，开展有针对性地进行加强学习。

为了进一步完善个性化学习推荐系统的设计与开发，将从以下两点进行完善：a)构建课程知识题库，为学习者筛选出合适的知识资源;b)自动匹配试题与知识点，节约知识点与试题绑定的时间成本。在后续研究中，将把个性化推荐系统的完善版本正式应用于普通中学进行教学辅导。

参考文献：

[1] 陈兵. 高职院校学生成绩管理系统的设计与实现 [D]. 重庆：重庆大学，2012. (Chen Bing. Design and implementation of student achievement management system in higher vocational colleges [D]. Chongqing: Chongqing University, 2012.)

[2] 韩冬梅. 无线网络在实施现代教育中的应用 [J]. 哈尔滨商业大学学报：自然科学版, 2010, 26 (4): 507-509. (Han Dongmei. Application of wireless network in implementing modern education [J]. Journal of Harbin University of Commerce: Natural Science Edition, 2010, 26 (4): 507-509.)

[3] 李志军. 数据挖掘技术在中职教学管理中的应用研究 [D]. 广州：广东技术师范学院，2017. (Li Zhijun. Application research of data mining technology in secondary vocational teaching management [D]. Guangzhou: Guangdong Technical Teachers College, 2017.)

[4] 潘燕. 关联规则下的数据挖掘算法分析 [J]. 信息记录材料, 2018, 19 (7):

212-213. (Pan Yan. Analysis of data mining algorithms under association rules [J]. Information Recording Materials, 2018, 19 (7): 212-213.)

[5] 李竟成, 孙晓青. 慕课对大学的影响研究 [J]. 南京工程学院学报: 社会科学版, 2018, 18 (02): 75-79. (Li Jingcheng, Sun Xiaoqing. Study on the Influence of MOOC on University [J]. Journal of Nanjing Institute of Technology: Social Science Edition, 2018, 18 (02): 75-79.)

[6] 牛玥. 我国在线教育平台发展模式及其问题对策探析——以网易公开课为例 [J]. 新媒体研究, 2018, 4 (10): 74-76. (Niu Yue. Analysis on the development model of online education platform in China and its countermeasures: taking netease open course as an example [J]. New Media Research, 2018, 4 (10): 74-76.)

[7] 刘润香. 中学生图形推理测验认知特征诊断及其与认知方式的关系研究 [D]. 南昌: 江西师范大学, 2007. (Liu Runxiang. Cognitive feature diagnosis of middle school students'graphic reasoning test and its relationship with cognitive style [D]. Nanchang: Jiangxi Normal University, 2007.)

[8] 朱天宇. 结合认知诊断方法的学生建模与应用研究 [D]. 合肥: 中国科学技术大学, 2018. (Zhu Tianyu. Student modeling and application research based on cognitive diagnosis method [D]. Hefei: University of Science and Technology of China, 2018.)

[9] 姜盼. 基于协同过滤的习题推荐系统设计与实现 [D]. 荆州: 长江大学, 2017. (Jiang Pan. Design and implementation of problem recommendation system based on collaborative filtering [D]. Jingzhou: Yangtze University, 2017.)

[10] 吴玺煜, 陈启买, 刘海, 等. 基于知识图谱表示学习的协同过滤推荐算法 [J]. 计算机工程, 2018, 44 (2): 226-232, 263. (Wu Xiyu, Chen Qimai, Liu Hai, *et al.* Collaborative filtering recommendation algorithm based on knowledge mapping representation learning [J]. Computer Engineering, 2018, 44 (02): 226-232, 263.)

[11] 李全, 刘兴红, 许新华, 等. 基于联合概率矩阵分解的个性化试题推荐方法 [J]. 计算机应用, 2018, 38 (3): 639-643, 649. (Li Quan, Liu Xinghong, Xu Xinhua, *et al.* Recommendation method for personalized test questions based on joint probability matrix decomposition [J]. Journal of Computer Applications, 2018, 38 (03): 639-643, 649.)

[12] 蒋昌猛. 基于知识点的个性化习题推荐研究 [D]. 西安: 西北大学, 2017. (Jiang Changmeng. Research on individualized exercises based on knowledge points [D]. Xi'an: Northwestern University, 2017.)

[13] 王雅婷. 数据挖掘在学生成绩分析中的应用 [J]. 现代信息科技, 2017, 1 (3): 96-97. (Wang Yating. Application of data mining in student performance analysis [J]. Modern Information Technology, 2017, 1 (3): 96-97.)

[14] 胡庆. 基于决策树的试卷知识点掌握程度分析研究 [D]. 南昌: 江西财经大学, 2014. (Hu Qing. Research on the degree of mastery of test paper knowledge based on decision tree [D]. Nanchang: Jiangxi University of Finance and Economics, 2014.)

[15] Jimmy D L T. DINA model and parameter estimation: A didactic [J]. Journal of Educational & Behavioral Statistics, 2009, 34 (1): 115-130.

[16] Shepp L A, Vardi Y. Maximum likelihood estimation for emission tomography [J]. Annals of the Institute of Statistical Mathematics, 2010, mi-1 (1): 113-121.

[17] 柳青. Linux 应用教程 [M]. 北京: 清华大学出版社, 2008. (Liu Qing. Linux Application Tutorial [M]. Beijing: Tsinghua University Press, 2008.)

[18] Hetland M L. Python 基础教程 [M]. 袁国忠, 译. 3 版. 北京: 人民邮电出版社, 2018. (Hetland M L. Python basic tutorial [M]. Yuan Guozhong, translation. 3rd ed. Beijing: People's Posts and Telecommunications Press, 2018.)